ST 540 hw1

Robin Baldeo

January 11, 2019

library(purrr)  
library(tidyr)  
library(knitr)

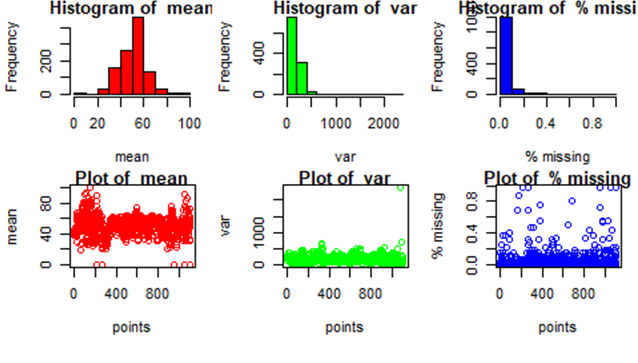
### (Part A)

#import file from desktop  
ozone<- read.csv("C:/Users/prl90/Documents/st540/ozone.csv", header=T)  
  
#flatten data frame  
data<- ozone[2:ncol(ozone)]%>%unlist  
  
#statistics  
data.frame("Mean" = mean(data, na.rm = T), "Sd" = sd(data, na.rm = T), "PerMissing" = sum(1 \* is.na(data))/length(data))%>%kable

|  |  |  |
| --- | --- | --- |
| Mean | Sd | PerMissing |
| 51.27333 | 17.26207 | 0.0432246 |

### (Part B)

#value extracted from the columns storage vector  
r <- nrow(ozone)  
c <- ncol(ozone)  
  
#mplot = vector for mean, sPlot = vector for var, pPlot = percent missing  
mPlot<- rep(0, r); sPlot<- rep(0, r); pPlot<- rep(0,r)  
  
for(i in 1:r){  
 v=rep(0, c-1)  
 for(j in 2:c){  
 v[j - 1] = ozone[i,j]  
 if(j == c){  
 mPlot[i] = ifelse(is.nan(mean(v, na.rm = T)),0,mean(v, na.rm = T) )  
 sPlot[i] = ifelse(is.nan(var(v, na.rm = T)) ,0,var(v, na.rm = T))   
 pPlot[i] = (sum(1 \* is.na(v))/c)  
 }  
 }  
}  
  
#plotting the histograms  
da<- list(list(mPlot, sPlot, pPlot),list("red", "green", "blue"),list("mean", "var", "% missing"))  
op <- par(pty="m", mfrow=c(3, 3), mar=c(4.2, 4.2, 1, 1))  
da%>% pwalk(~hist(x = ..1, col =..2, xlab = ..3 , main = paste("Histogram of ", ..3)))  
da%>%pwalk(~plot(..1, col = ..2, ylab = ..3, xlab = "points", pch = 1, main = paste("Plot of ", ..3) ))  
par(op)



### (Part C)

model<- lm(mPlot~sPlot + pPlot)  
summary(model)

##   
## Call:  
## lm(formula = mPlot ~ sPlot + pPlot)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31.054 -7.485 0.740 6.345 44.981   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 48.589550 0.632819 76.783 < 2e-16 \*\*\*  
## sPlot 0.017099 0.002962 5.773 1.01e-08 \*\*\*  
## pPlot -10.880774 3.728495 -2.918 0.00359 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.18 on 1097 degrees of freedom  
## (6 observations deleted due to missingness)  
## Multiple R-squared: 0.03674, Adjusted R-squared: 0.03498   
## F-statistic: 20.92 on 2 and 1097 DF, p-value: 1.212e-09